CHEMICAL ENGINEERING TRANSACTIONS Volume 24, 2011
Editor Sauro Pierucci
 ISBN 978-88-95608-15-0 ISSN 1974-9791
DOI: 10.3303/CET1124055

An Homotopy Method for Global Optimization of Continuous Models

Karim Alloula[1*], Jean-Pierre Belaud[1], Jean-Marc Le Lann[1]

[1] Université de Toulouse, Laboratoire de Génie Chimique (CNRS UMR 5503), INPT-ENSIACET
4, allée Emile Monso, BP 44362, 31030 Toulouse Cedex 4, France
*Karim.Alloula@ensiacet.fr

An original approach to global optimization of continuous models is introduced. It belongs to the class of homotopy continuation methods, but "only" requires non linear equation systems to be solved. Unconstrained and non-linearly constrained optimization problems are specified nearly the same way. They are solved by coupling a robust Newton formulation for under determinate systems and a heuristic estimating the global minimum value by means of the discrete Legendre-Fenchel biconjugate of the criterion. For the time being, the main drawback of the method is the too important number of function evaluations near by the global minimum. However, its success rate being very good on test problems, such as the global optimization of Lennard-Jones atomic clusters, it should be investigated further.

1. Introduction

Among the main classes of global optimization methods, reviewed for example in Floudas and Gounaris (2009), homotopy continuation methods remain quite confidential within the CAPE community. One reason is the difficulty to handle several optimization problems starting from a "simple" one, and moving to the problem of interest.

This paper introduces an homotopy method, suitable for continuous models only, which solves global optimization problems under non linear constraints. This original approach may be attractive because, instead of tackling intermediate optimization problems, it requires "only" to apply a Newton iterative scheme to non linear equation systems.

2. Unconstrained global optimization

Let illustrate the method principle with the unconstrained global optimization of a continuous criterion $f(x)$, where $x \in \mathbb{R}^m$. We assume that we know some global minimum underestimate $\bar{\lambda}$ of the criterion. Instead of solving the initial optimization problem $min_{x\in\mathbb{R}^m} f(x)$, we are interested in a parameterized non linear problem (P_i): $g_{\lambda_i}(x) = f(x) - \lambda_i = 0$. For a given value of λ_i, we compute a sequence of K Newton iterates $x_{i,1}$, $x_{i,2} = \mathrm{N}_{g_{\lambda_i}}(x_{i,1}), x_{i,3} = \mathrm{N}_{g_{\lambda_i}}(x_{i,2}), \ldots$ where $\mathrm{N}_{g_{\lambda_i}}$ is a Newton operator

Please cite this article as: Alloula K., Belaud J.P. and Le Lann J.M., 2011, An homotopy method for global optimization of continuous models, Chemical Engineering Transactions, 24, 325-330
DOI: 10.3303/CET1124055

associated to g_{λ_i}. Among those Newton iterates, we consider only a subset $S_i = \{j \in \{1,2,\dots,K\}; g_{\lambda_i}(x_{i,j}) \leq 0\}$. At any Newton iterate $x_{i,j}$, where j belongs to S_i, the criterion value is less than λ_i.

The principle is to build a strictly decreasing sequence of criterion values from S_i subsets. First of all, we try to exhibit some $x_{0,1}$ verifying $\|Df(x_{0,1})\|_\infty \geq \varepsilon$. We define $\lambda_0 = f(x_{0,1})$ and $S_0 = \{1\}$. The next subsets are built according to the following:

- If S_i is not empty, pick an iterate x_{i,j^*} from S_i according to some strategy. $f(x_{i,j^*}) = \lambda_i^* \leq \lambda_i$.
  - If $\|Df(x_{i,j^*})\|_\infty < \varepsilon$ then consider that $x_i^* = x_{i,j^*}$ approximates a global minimum of f;
  - If $\|Df(x_{i,j^*})\|_\infty \geq \varepsilon$ then, from $x_{i+1,0} = x_{i,j^*}$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = g_{\bar{\lambda}}(x) = f(x) - \bar{\lambda} = 0$ to produce S_{i+1};
- If S_i is empty, consider the last non empty set S_a and define $\lambda_{i+1} = \frac{\lambda_a^* + \lambda_i}{2}$.
  - If $|\lambda_{i+1} - \lambda_a^*| < \varepsilon$ then consider that x_a^* approximates a global minimum of f;
  - If $|\lambda_{i+1} - \lambda_a^*| \geq \varepsilon$ then, from $x_{i+1,0} = x_a^*$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = f(x) - \lambda_{i+1} = 0$ to produce S_{i+1}.

3. Non-linearly constrained global optimization

The previous method can be easily extended to non-linearly constrained global minimization. Equality constraints are added to non linear problems (P_i). Inequality constraints are transformed into equality constraints by introducing slack variables. Then those equality constraints are added to (P_i) problems.

Once again we assume that we know some global minimum underestimate $\bar{\lambda}$ of the criterion. Instead of solving the initial optimization problem $min_{x\in\mathbb{R}^m} f(x)$ under the constraints $c(x) = 0$, we are interested in a parameterized non linear problem (P_i): $g_{\lambda_i}(x) = \begin{pmatrix} f(x) - \lambda_i \\ c(x) \end{pmatrix} = 0$. For a given value of λ_i, we compute a sequence of K Newton iterates $x_{i,1}$, $x_{i,2} = \mathrm{N}_{g_{\lambda_i}}(x_{i,1}), x_{i,3} = \mathrm{N}_{g_{\lambda_i}}(x_{i,2}), \dots$ where $\mathrm{N}_{g_{\lambda_i}}$ is a Newton operator associated to g_{λ_i}. Among those Newton iterates, we consider only a subset $S_i = \{j \in \{1,2,\dots,K\}; f(x_{i,j}) - \lambda_i \leq 0 \wedge c(x_{i,j}) = 0\}$. At any Newton iterate $x_{i,j}$, where j belongs to S_i, the criterion value is less than λ_i.
The principle is to build a strictly decreasing sequence of criterion values from S_i subsets. First of all, we try to exhibit some $x_{0,1}$ verifying $\|Df(x_{0,1})\|_\infty \geq \varepsilon \wedge c(x_{0,1}) = 0$. We define $\lambda_0 = f(x_{0,1})$ and $S_0 = \{1\}$. The next subsets are built according to the following:

- If S_i is not empty, pick an iterate x_{i,j^*} from S_i according to some strategy. $f(x_{i,j^*}) = \lambda_i^* \leq \lambda_i$.
  - If $\|Df(x_{i,j^*})\|_\infty < \varepsilon$ then consider that $x_i^* = x_{i,j^*}$ approximates a global minimum of f;
  - If $\|Df(x_{i,j^*})\|_\infty \geq \varepsilon$ then, from $x_{i+1,0} = x_{i,j^*}$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = g_{\bar{\lambda}}(x) = \begin{pmatrix} f(x) - \bar{\lambda} \\ c(x) \end{pmatrix} = 0$ to produce S_{i+1};
- If S_i is empty, consider the last non empty set S_a and define $\lambda_{i+1} = \frac{\lambda_a^* + \lambda_i}{2}$.
  - If $|\lambda_{i+1} - \lambda_a^*| < \varepsilon$ then consider that x_a^* approximates a global minimum of f;
  - If $|\lambda_{i+1} - \lambda_a^*| \geq \varepsilon$ then, from $x_{i+1,0} = x_a^*$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = \begin{pmatrix} f(x) - \lambda_{i+1} \\ c(x) \end{pmatrix} = 0$ to produce S_{i+1}.

To summarize, the non linearly constrained global optimization algorithm differs from the unconstrained global optimization procedure only in the following points:

1. An initial guess $x_{0,1}$ has to be found on the constraints variety;
2. Homotopy functions g_{λ_i} incorporate the residuals associated to the equality constraints. g_{λ_i} is a vectorial function, defined from $\mathbb{R}^m$ to $\mathbb{R}^n$, where $n < m$ in the usual case (criterion value is optimized within a constraints variety not restricted to a single point). $Dg_{\lambda_i}(x)$, derivative of g_{λ_i} at x, is an n by m matrix;
3. S_i subsets definition is modified to take constraints validation into account. One should notice that any of the selected points in the S_i subsets is a feasible point.

4. Numerical challenges and answers

The homotopy method introduced here seems to be easier to implement than other homotopy continuation methods. First, it can be viewed as a Newton homotopy, which is simple and efficient (Yakoubsohn, 2003). Second, instead of tackling intermediate optimization problems, it requires "only" to apply a Newton iterative scheme to a family of non linear equation systems. However, some numerical and software challenges have to be tackled.

4.1 Under determinate non linear systems

The first numerical challenge comes from the fact that the non linear systems are under determinate, with more variables than equations. Consequently, a generalized Newton formulation, adopted from Dedieu (2006), is required. Assuming F is the residual function to nullify, the generalized Newton operator associated to F in the surjective

case is defined by $N_F(u) = u - [DF(u)]^{\dagger}.F(u)$, where $[DF(u)]^{\dagger}$ stands for the Moore-Penrose pseudo-inverse of the Jacobian matrix of F at u.

When F takes its value in $\mathbb{R}$ (unconstrained global optimization), the generalized Newton operator associated to F is $N_F(u) = u - \frac{F(u)}{DF(u) \cdot DF(u)}.DF(u)$.

4.2 Global minimum underestimate

The second numerical challenge is related to finding a good underestimate of the global minimum. Obviously, most of the time, a good approximation of the global minimum is not available when starting the method. So, instead of using a single global minimum underestimate $\bar{\lambda}$, as stated in parts 2 and 3, a first refinement is to use a sequence of global minimum underestimates $(\bar{\lambda}_i)_i$, each one being predicted from the criterion values already computed $(\lambda_j{}^*)_j$. When $\bar{\lambda}_i$ prediction is based only on the last two criterion values $\lambda_i{}^*$ and $\lambda_{i-}{}^*$, a strategy may be to try to decrease the criterion value twice more than previously done. In such a case, $\bar{\lambda}_i = \lambda_i{}^* - 2(\lambda_{i-}{}^* - \lambda_i{}^*)$. To initiate the process, $\bar{\lambda}_0$ is set to a numerical value representing $-\infty$. Assuming S_1 is not empty, $\bar{\lambda}_1$ is set to $\bar{\lambda}_1 = \lambda_1{}^* - 2(\lambda_0 - \lambda_1{}^*)$.

Unfortunately, this first refinement is not always efficient for leaving the basins of attraction of local minima. We have to incorporate some global knowledge to the previous prediction which is based only on the local behavior of the criterion. So, the second refinement consists in estimating a convex hull of the criterion by applying twice the Legendre-Fenchel transform to it. Calculating the Legendre-Fenchel bi-conjugate from the analytical expression of the criterion may be harder than the initial global optimization problem: instead of one optimization problem, one has to solve two parameterized optimization problems! In fact, the bi-conjugate of the criterion function is estimated only at points where the criterion has already been calculated. Moreover, this estimation is based on a discrete Legendre-Fenchel transform. Instead of calculating the conjugate f^* of a real function f from $f^*(k) = \sup_x \{k \cdot x - f(x)\}$, we assume that $f^*(x_i) \simeq \max_j \{x_i \cdot x_j - f(x_j)\}$, where $\{f(x_1), f(x_2), \ldots\}$ are previous criterion evaluations. The first prediction formula $\bar{\lambda}_i = 2(\lambda_{i-}{}^* - \lambda_i{}^*)$ is replaced by:
$\bar{\lambda}_i = min\left[2(\lambda_{i-}{}^* - \lambda_i{}^*), \min_i f^{**}(x_i)\right]$.

Such a prediction leads to an underestimation of any minimum, either local or global. For local minima, this strategy is adequate. But, when reaching the basin of attraction of the global minimum, this strategy leads to unnecessary criterion evaluations because we are looking for points $x_{i,j}$ satisfying $f(x_{i,j}) - \lambda_i \leq 0$ without success, until λ_i becomes slightly greater than the global minimum.

4.3 Symbolic numeric calculations

From the software point of view the challenge is, starting from the initial optimization problem $min_{x \in \mathbb{R}^m} f(x)$:

- To obtain automatically the g_{λ_i} functions and their analytical derivatives Dg_{λ_i};
- To obtain the numerical values of $\left[Dg_{\lambda_i}(u)\right]^{\dagger}$ matrices;
- To obtain a numerical approximation of the Legendre-Fenchel bi-conjugate ${g_{\lambda_i}}^{**}$ from some values of g_{λ_i}.

Those symbolic processing steps and those numerical evaluations both take place within *eXMSL*, a symbolic and numerical calculation system (Alloula et al., 2009). Very accurate Jacobian matrices are obtained automatically from the model. The IMSL® numerical library is in charge of the Moore-Penrose pseudo-inverse calculations. The Legendre-Fenchel bi-conjugates are evaluated without any care about CPU time.

5. Application to global optimization of Lennard-Jones atomic clusters

The global optimization of Lennard-Jones atomic clusters is a classical test problem for global optimization methods. The problem description can be found in Daven (1996). The strategy described in this paper, and coded within *eXMSL*, was applied to clusters up to 100 atoms. The initial guess was always randomly selected without any a priori knowledge, making the problems harder to solve.

With an unconstrained formulation of the problem, we retrieved the best criterion values given within the literature only when the number of atoms was less than 20. The success rate decreases with N, being 80% for $N = 6$, and vanishing to 0% around $N = 15$.

In order to check the method validity on non-linearly constrained problems, we stated that all the inter-particle distances were greater than 0.6187 (Vinko, 2005). For a cluster of N atoms, this results in $N(N-1)/2$ constraints. With this constrained formulation of the problem, we retrieved all the best criterion values given within the literature for $3 \leq N \leq 100$. The success rate increases with N, being 70% for $N = 6$, and reaching 100% when $N \geq 30$. Combined with the equation $f(x) - \lambda = 0$, constraints on the inter-particle distances provide a good path for the homotopy method to reach a global minimum. When N is quite small, the feasible domain defined by "the inter-particle distances are to be greater than 0.6187" is important when compared to the whole variable space (for $N = 6$, the inter-particle distances at the global minimizer are close to 0.995531). When $N \geq 30$, the feasible domain defined by "the inter-particle distances are to be greater than 0.6187" is small when compared to the whole variable space. Then, the main difficulty becomes to find an initial point belonging to the constraint variety. This problem consists in solving an under determinate non linear system of $1 + N(N-1)/2$ equations and $3N + N(N-1)/2$ variables. The generalized Newton formulation, cited previously, was applied successfully, even for initial guesses without any physical meaning.

6. Conclusion and further work

This paper introduces a deterministic method for global optimization. Applied to non-linearly constrained continuous models, this homotopy method produces a feasible path along which the criterion value decreases. Points of the homotopy curve come from a Newton process where a generalized Newton operator tries to keep iterates on the constraint variety. Two benefits are associated with such a feasible path optimization:

- the variables remain in the validity domain of the (thermodynamic) models;
- the search space being limited to the constraint variety, the probability for the algorithm to reach and stay at a non global minimum is reduced.

This method was applied successfully to a collection of test problems. Unfortunately, for the time being, the number of function evaluations remains important, mainly because of a costly bisection process: the homotopy parameter λ_i is successively increased or decreased, depending on whether some $x_{i,j}$ verifying $f(x_{i,j}) \leq \lambda_i$ is found or not. For the method to be not only robust, but also efficient, the bisection strategy has to be improved. Two ways are investigated:

- study the discrete Legendre-Fenchel transform which, when applied twice to a set of points on the criterion representative surface, provides us a convex hull of those points. The closer this convex hull is to the criterion representative surface, the better is the prediction of the next homotopy parameter value;
- find the "best" bisection parameter value to be applied when some $x_{i,j}$ verifying $f(x_{i,j}) \leq \lambda_i$ is not found. Until now, this bisection parameter is set to 0.5. Its value may be related to the absolute error $|f(x_{i,j}) - \lambda_i|$.

This homotopy method replaces the initial optimization problem by a sequence of non linear systems involving continuous variables. We would like to extend such an approach to mixed integer non linear programming, discrete variables being considered as continuous ones by adding constraints in order to restrict their allowed values.

References

Alloula K., Belaud J.-P. and Le Lann J.-M., 2009, A co-operative model combining computer algebra and numerical calculation for simulation, Computer-Aided Chemical Engineering, 26, 889-894.

Daven D.M., Tit N., Morris J.R. and Ho K.M., 1996, Structural optimization of Lennard-Jones clusters by a genetic algorithm, Chemical Physics Letters, 256, n°1-2, 195-200.

Dedieu J.-P., 2006, Points fixes, zéros et la méthode de Newton, Springer Berlin Heidelberg.

Floudas C. A., Gounaris C. E., 2009, A review of recent advances in global optimization, Journal of Global Optimization, 45, n°1, 3-38.

Vinko T., 2005, Minimal inter-particle distance in atom clusters, Acta Cybernetica, 0, 1-15.

Yakoubsohn J.-C., 2003, Numerical Elimination, Newton Method and Multiple Roots, Algorithms Seminar 2001–2002, F. Chyzak (ed.), INRIA, 49–54.